

Julkaisuportaali: tilannekatsaus ja alustava suunnitelma vuoden 2013 toiminnasta

1. Yhteenveto

Kansalliskirjasto rakensi syksyn 2012 aikana prototyypin avoimen lähdekoodin VuFind-ohjelmistoon perustuvasta Julkaisuportaalista. Projektin tavoitteena oli luoda suomalaisten yliopistojen julkaisutietojen selailuun ja niistä tehtäviin hakuihin soveltuva käyttöliittymä, joka lisäisi OKM:n tekemän julkaisutietojen keruun ja samalla myös julkaisutietoja hyödyntävän yliopistojen uuden rahoitusmallin läpinäkyvyyttä.

Prototyypissä käytettiin testiaineistona opetus- ja kulttuuriministeriön yliopistoilta keräämiä vuoden 2011 julkaisutietoja, jotka konvertoitiin Excel-tiedostoista soveltuvaan formaattiin ja importoitiin eräajona prototyyppiä varten perustettuun VuFind-instanssiin. VuFindin käyttöliittymän selailuvaihtoehdot ja toiminnallisuudet pyrittiin räätälöimään julkaisutietojen kannalta optimaalisiksi. Kolmen henkilötyökuukauden projektissa saavutetut tulokset osoittavat, että VuFind ohjelmistona soveltuu hyvin julkaisutietojen esittämiseen.

Vaikka prototyyppi on teknisten ratkaisujen osalta avattavissa laajempaan käyttöön suhteellisen vähäisellä työllä, sen konseptista ja käyttötarkoituksista kaivataan vielä palautetta yliopistoilta ja muilta keskeisiltä sidosryhmiltä. Tähän mennessä toteutetussa pilottiprojektissa keskityttiin toiminnallisuuksien luomiseen, eli palvelun brändääminen ja vastuutahojen määrittely edellyttävät vielä työtä. Alustavasti on päätetty, että palvelu viimeistellään julkaisukuntoon huhtikuuhun 2013 mennessä.

Vuoden 2013 aikana portaaliin lisätään ainakin yliopistojen vuoden 2012 julkaisutiedot, jotka saadaan tämänhetkisen käsityksen mukaan samantapaisessa formaatissa kuin vuoden 2011 tiedot. Lisäksi on käyty alustavaa keskustelua julkaisujen tiedonkeruuseen liittyvien prosessien kehittämisestä ja automatisoinnista. Mahdollisen rajapintojen ja prosessien kehittämisen aikataulu tai sen julkaisuportaalin osalta aiheuttama työmäärä ei kuitenkaan ole vielä tiedossa.

2. Projektin toiminta ja tulokset

2.1. Projektin eteneminen

Kansalliskirjasto sai Julkaisuportaalin prototyypin rakentamista varten kolmen henkilötyökuukauden työpanosta vastaavan rahoituksen CSC:n koordinoimasta RAKETTI-hankkeesta. Prototyyppiin liittyvä työ käynnistyi syyskuussa 2012 ja jatkui tammikuun 2013 alkuun. Projektia varten tehtyyn työhön osallistuivat Kansalliskirjastossa tietojärjestelmäasiantuntijat Nicholas Volk ja Jyrki Ilva.

Projektin etenemisestä järjestettiin kokous Kansalliskirjaston, OKM:n, CSC:n ja TSV:n edustajien kesken 15.11.2012. Lisäksi projektia esiteltiin mm. 4.10. Tampereen teknillisellä yliopistolla järjestetyssä yliopistojen tutkimustietojärjestelmiin liittyneessä tapaamisessa, 11.12. RAKETTI-TUTKI:n ohjausryhmässä ja 13.12. Julkaisufoorumin ohjausryhmässä.

2.2. Julkaisutietojen käsittely

Prototyypin testiaineistoksi saatiin elo-syyskuussa CSC:ltä (Yrjö Leino) OKM:n yliopistoilta keräämät vuoden 2011 julkaisutiedot. Viitetiedot saatiin kolmena erillisenä Excel-tiedostona, joiden sisältämissä kentissä oli jonkin verran eroavaisuuksia. Exceleissä oli tunnistettu yliopistojen yhteisjulkaisut, mutta eri yliopistojen samasta julkaisusta raportoimat tiedot olivat eri riveillä. Samoin Exceleiden mukana saatiin julkaisujen alustavat julkaisuforumiluokitukset.

VuFind-importointia varten Excel-muotoiset viitetiedot konvertoitiin Dublin Core -pohjaiseen metadataformaattiin, jota oli laajennettu julkaisutietojen kannalta tarpeellisilla lisäkentillä. Konversion yhteydessä dataa muokattiin ja sen pohjalta generoitiin myös uusia tietoja, mm. yhdistämällä tekijöiden nimet raportoiviin yliopistoihin ja lisäämällä yliopiston nimi kunkin julkaisuun osallistuneen yliopiston raportoimien sisäisten affiliaatioiden alkuun.

Yliopistojen raportoimissa tiedoissa oli oletusten mukaisesti runsaasti puutteita, epä johdonmukaisuuksia ja virheellisiä tietoja, joita pyrittiin mahdollisuuksien mukaan korjaamaan ja normalisoimaan ennen konversiota. Esim. tekijöiden nimiä korjattiin ohjelmallisesti oikeaan formaattiin (sukunimi, etunimi) niiltä osin kuin oikean muodon tunnistaminen automaattisesti oli mahdollista. Samoin yhteisjulkaisuja koskevien tietueiden yhdistäminen edellytti valintoja sen suhteen minkä yliopiston raportoimista tiedoista mikin kenttä poimittiin. Esim. otsikkojen kohdalla päädyttiin yksinkertaisesti käyttämään pisintä versiota, koska sen saattoi olettaa useimmissa tapauksissa sisältävän eniten informaatiota.

Konversiota muokattiin samanaikaisesti muun kehitystyön kanssa, jolloin kenttien formaattia voitiin optimoida käyttöliittymän toiminnallisuuden mukaan. Käytännössä tämä tarkoitti sitä, että aina konversioskripteihin tehtyjen muutosten jälkeen julkaisutiedot konvertoitiin ja ajettiin uudelleen VuFindiin.

Myös merkistöihin kiinnitettiin konversiossa erityishuomiota, mutta esim. julkaisudatan sisältämä kyriliikka oli korruptoitunut jo Excel-vaiheessa tai ennen sitä, eikä sitä ollut enää mahdollista korjata automaattisesti. Tämä ongelma koski noin 60 julkaisua, ja sen korjaaminen edellyttäisi tietojen poimimista yksitellen yliopistojen omista järjestelmistä.

Portaalin sisältämiä tietoja on toisaalta mahdollista päivittää eräajona myös myöhemmin, jos esim. julkaisujen julkaisuforumiluokituksiin tulee vielä korjauksiin. Tietojen päivittäminen edellyttää sitä, että korjatut luokitukset on yhdistetty CSC:n kullekin julkaisulle tietojen yhdistämisen yhteydessä antamaan tunnistenumeroon, jota myös julkaisuportaaliissa käytetään tietueiden tunnisteina.

2.3. Portaalin prototyyppi

Prototyyppiä varten luotiin Kansalliskirjaston Helsingin yliopiston tietotekniikkakeskuksen konesalissa sijaitsevaan palvelininfrastruktuuriin oma virtuaalipalvelin, johon asennettiin VuFind-ohjelmiston tuorein vakaa versio (1.3.). Tämän testikäyttöön tarkoitetun VuFind-instanssin verkko-osoite on

<http://iupo-kktest.lib.helsinki.fi/vufind>

Instanssia koskevat palomuurirajaukset poistettiin testaamisen ja prototyypin esittelemisen helpottamiseksi marraskuun alussa, minkä jälkeen siihen on ollut mahdollista päästä käsiksi myös Helsingin yliopiston paikallisverkon ulkopuolelta. Käyttöliittymän indeksoituminen Internetin hakukoneisiin on kuitenkin toistaiseksi estetty.

Ainakaan testikäytössä järjestelmässä ei ole ilmennyt minkäänlaisia kapasiteettiongelmia. VuFind on tarkoitettu hyvinkin laajojen data-aineistojen käyttöliittymäksi, ja virtuaalipalvelimelle voidaan myös tarvittaessa antaa joustavasti lisää resursseja, eli ei ole todennäköistä, että merkittäviä kapasiteettiongelmia ilmenisi ainakaan nykyisen toimintamallin puitteissa.

2.4. Toteutetut ominaisuudet

Julkaisuportaalin prototyyppissä pyrittiin hyödyntämään VuFind-ohjelmiston olemassaolevia ominaisuuksia, eli siinä ei tehty laajaa uudelleenkoodaamista. VuFindin käyttöliittymä ja sen taustalla oleva Apachen Solr-teknologia tarjoaa kuitenkin varsin suuria räätälöintimahdollisuuksia, joiden avulla käyttöliittymä pyrittiin optimoimaan nimenomaan julkaisutietojen esittämisen tarpeisiin. Tämä edellytti myös joidenkin laajennusten tekemistä VuFindin tietomalliin.

Excel-datasta tuotettujen julkaisukohtaisten Dublin Core -tiedostojen kentät mapattiin VuFindin käyttämään tietomalliin siten, että niiden sisältämiä tietoja voidaan käyttää järkevästi VuFindin tietuenäytössä, faseteissa ja hakuindekseissä.

Tavoitteena oli se, että dataa on käyttöliittymässä mahdollista tarkastella monesta eri näkökulmasta. Käyttöliittymän sivupalkissa näkyviä valmiita rajausmahdollisuuksia (fasetteja) prototyyppissä on kahdeksan: organisaatio, tekijän nimi, koulutusala, tieteenala, julkaisufoorumiluokka, julkaisutyyppi, kieli, ja julkaisumaa.¹

Rajaa hakua

Organisaatio

[Helsingin yliopisto](#) (10153)

[Turun yliopisto](#) (4001)

[Aalto-yliopisto](#) (3624)

[Itä-Suomen yliopisto](#) (2862)

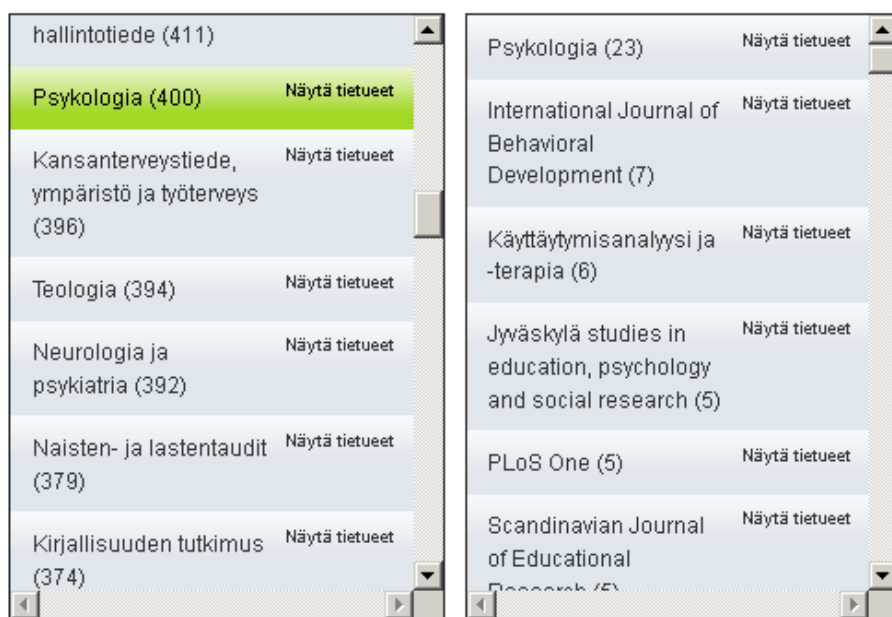
[Jyväskylän yliopisto](#) (2788)

[lisää ...](#)

Fasettien tarjoamien rajausmahdollisuuksien lisäksi tietuenäytössä näkyvistä tiedoista tehtiin linkkejä, joiden kautta on mahdollista päästä selaamaan muita julkaisuja, joissa esiintyy sama tieto (esim. tekijän nimi, lehden nimi, kustantajan nimi, ISSN, tieteenala tai paikallinen affiliaatio). VuFindin tarkennettuun hakuun rakennettiin mahdollisuus valita samalla kertaa (ctrl-näppäimellä ja hiirellä) useampia vaihtoehtoja samasta kategoriasta. Lisäksi erilliselle listaussivulle ("Sela luettelo"-linkki

¹ Julkaisuvuosi jätettiin toistaiseksi pois rajausmahdollisuuksista, koska mukana on dataa vain yhdeltä vuodelta.

käyttöliittymän alalaidassa) toteutettiin lisää lajittelumahdollisuuksia, joilla voi selata esim. suosituimpia lehtiä tai kustantajia tieteenaloittain tai organisaatioittain.



Etenkin tekijätietojen ja tieteenala/koulutuslaluokitusten kohdalla jouduttiin tekemään valintoja sen suhteen, mitä tietoja eri yhteyksissä käytetään. VuFindin tietomalli nosti julkaisun ensimmäisen tekijän erityisrooliin, mikä oli kuitenkin julkaisudatan kannalta ongelmallista, eli tietomallia muutettiin tältä osin. Lisäksi data sisälsi kolmenlaisia tekijätietoja: julkaisun täydellinen tekijälistaus (tai 20 ensimmäistä tekijää, jos tekijöitä oli enemmän), organisaatioiden raportoimat tekijät ja tekijä-organisaatio-parit. Ainakin tässä vaiheessa päädyttiin siihen, että faseteissa käytetään yliopistojen raportoimien tekijöiden nimiä, mutta hakutuloksissa ja tietuenäytössä näytetään kuitenkin täydelliset tekijätiedot. Tietuenäytössä näytetään lisäksi tekijä-organisaatio-parit.

Tieteenala

[Tietojenkäsittely ja informaatiotieteet](#) (2134)

[Fysiikka](#) (2044)

[Liiketaloustiede](#) (1700)

[Kasvatustieteet](#) (1556)

[Historia ja arkeologia](#) (1386)

[lisää ...](#)

Tieteenalojen ja koulutusalojen osalta päädyttiin siihen, että faseteissa käytetään vain yliopistojen ensisijaisina raportoimia aloja. Yliopistot olivat määritelleet etenkin tieteenaloja vaihtelevilla kriteereillä (julkaisun tieteenalan sijasta esim. laitostietojen pohjalta), ja osa yliopistoista oli raportoinut kaikille julkaisuille säännönmukaisesti monia tieteenaloja. Vaikka ensimmäiseksi määritelty tieteenala ei todennäköisesti ole kaikissa tapauksissa osuvin, toissijaiset alat olisivat kuitenkin useimmissa tapauksissa vain sekoittaneet luokituksia lisää.

Käyttöliittymää suunniteltiin siten, että se olisi käytettävissä kolmella kielellä (suomi, ruotsi ja englanti). Tavoitteena oli pyrkiä siihen, että käyttöliittymän otsikoiden ja ohjetekstien lisäksi myös organisaatioiden nimet, koulutusalat, tieteenalat, julkaisutyyppit, kielet ja julkaisumaa (Suomi/muu) näkyisivät käyttöliittymässä erikielisinä kielivalinnan mukaan. Täydellinen monikielisyys ei kuitenkaan ole mahdollista, koska datassa itsessään on sellaista suomen- tai ruotsinkielistä tietoa (esim. paikalliset affiliaatiot), jonka kieliversiointi ei ole mahdollista. Käytännössä monikielisyyden

toteuttaminen jäi vielä prototyypivaiheessa keskeneräiseksi, osin siksi, ettei kaikista käännettäviksi suunnitelluista tiedoista ollut saatavilla valmiita käännöksiä kaikilla kolmella kielellä. Lisäksi kehitystyön aikana ilmeni, että monikielisyyden tekninen toteutus saattaa aiheuttaa jonkin verran ongelmia näitä tietoja koskeville hauille.

Julkaisuportaalin käyttöliittymästä pyrittiin selkeyden vuoksi karsimaan pois kaikki tarpeettomat ominaisuudet. Muun muassa VuFindin tarjoama kirjautumismahdollisuus poistettiin käytöstä, sillä käyttäjätileistä ei nähty olevan Julkaisuportaalin todennäköisten käyttötarkoitusten kannalta oleellista hyötyä.

2.5. Prototyypivaiheen tulokset

Projektin aikana selvisi varsin nopeasti, että VuFindin käyttöliittymä soveltuu hyvin yliopistojen julkaisutietojen esittämiseen. Huolimatta siitä, että julkaisudatan käsittely ja muokkaaminen vaativat enemmän työtä kuin projektia suunniteltaessa uskottiin, käyttöliittymään pystyttiin toteuttamaan sekä kaikki etukäteen suunnitellut ominaisuudet että useita sellaisia toiminnallisuuksia, jotka suunniteltiin vasta projektin aikana.

Projektissa tuotettu prototyyppi on jo sinällään täysin käyttökelpoinen, ja vaikka se vaatii vielä jonkin verran viimeistelyä ja brändäämistä, se on kuitenkin mahdollista saattaa julkaisukuntoon melko nopeasti. Projektissa määriteltiin myös julkaisuportaaliin syötettävälle datalle yhdenmukainen Dublin Core -pohjainen formaatti, jota on mahdollista käyttää yhtenä lähtökohtana sekä tulevien konversioiden että mahdollisten nykyistä dynaamisempien järjestelmien välisten rajapintojen suunnittelussa.

3. Käyttöönotto

Julkaisuportaali on alustavan suunnitelman mukaan tarkoitus avata julkiseen käyttöön huhtikuussa 2013. Tämä edellyttää päätöksiä portaalin brändämisestä, verkko-osoitteesta ja siihen liittyvistä vastuutahoista sekä sen teknisten ratkaisujen ja visuaalisen ulkoasun viimeistelemistä. Lisäksi palvelusta ja sen käyttötarkoituksista on oleellisen tärkeää keskustella yliopistojen ja muiden sidosryhmien kanssa.

3.1. Palautteen kerääminen yliopistoilta

Portaalia esitellään alkuvuoden 2013 aikana yhteistyössä OKM:n kanssa yliopistojen edustajille, ja siihen liittyen järjestetään rajatulle osallistujajoukolle suunnattu verkkokysely, jossa kerätään palautetta portaalin ominaisuuksista, siihen kohdistuvista odotuksista ja toiveista sekä arvioita sen mahdollisista käyttötarkoituksista. Yliopistojen kanssa on syytä keskustella myös datan omistajuudesta ja käyttöehdoista.

3.2. Portaalin kohderyhmät

Portaalin voi olettaa kiinnostavan tutkimushallinnon lisäksi myös laajempia käyttäjäryhmiä, mukaan lukien tutkijat, tiedetoimittajat ja muut yliopistojen julkaisutoiminnasta kiinnostuneet.

Ellei yliopistoilta saatavassa palautteessa nouse esiin uusia näkökohtia, portaali on todennäköisesti mielekästä avata suoraan mahdollisimman avoimeen ja laajaan käyttöön. Käyttöoikeuden rajaaminen edellyttäisi uusia teknisiä ratkaisuja ja vaatisi resurssien suuntaamista käyttöoikeuksien hallintaan, eli siitä aiheutuisi merkittäviä lisäkustannuksia ja se hankaloittaisi palvelun käyttämistä.

Vaikka portaalin sisältämät tiedot ovat joiltakin osin puutteellisia ja yksittäisten julkaisujen osalta kenties virheellisiäkin, sillä on kuitenkin arvoa nimenomaan laajana ja ongelmista huolimatta laadukkaana tietovarantona. Lisäksi tietojen julkisen näkyvyyden voi olettaa kannustavan yliopistojen omien julkaisutietoja koskevien tallennus- ja tarkistusprosessiensa kehittämiseen.

3.3. Palvelun brändin ja vastuutahojen määrittäminen

Ennen portaalin julkaisemista tarvitaan päätökset sen nimestä ja verkko-osoitteesta, eli mm. sijaitseeko se jonkin organisaation domain-nimen alla (esim. julkaisuportaali.organisaatio.fi) vai onko sillä kokonaan oma verkko-osoite (esim. www.julkaisuportaali.fi). Koska portaali on pohjimmiltaan vain käyttöliittymä yliopistojen julkaisudataan, sen visuaalinen ulkoasu voi todennäköisesti olla melko pelkistetty, eli päähuomio on nimenomaan datassa.

Palvelun teknisen ylläpidon osalta vastuuorganisaatio on Kansalliskirjasto, mutta koska julkaisutiedot ovat peräisin muista organisaatioista ja järjestelmistä, eikä niitä muokata portaalissa, mahdollisten sisältöä koskevien kysymysten ja kommenttien käsittelyyn täytyy määritellä erikseen omat prosessinsa ja vastuutahonsa.

3.4. Palvelun viimeistely ja avaaminen

Julkaisuportaalin ulkoasun muokkaaminen (ml. mahdollinen logo) pyritään toteuttamaan suhteellisen kevyellä organisaatiolla siten, että lopputulos on melko lähellä VuFindin ja kansallisen tiedonhakupalvelun Finnan perusulkoasua.

Portaalin käyttöliittymässä vielä olevat virheet ja epäjohtonmukaisuudet pyritään korjaamaan, ja käyttöliittymän kieliversiointi (suomi, ruotsi, englanti) toteutetaan loppuun asti. Portaalille laaditaan suppeahkot käyttöohjeet, joissa kerrotaan hakuohjeiden lisäksi myös perustiedot sen sisältämästä datasta. Palvelun yhteyteen lisätään vastuutahojen yhteystiedot ja palautelomake sitä koskevien kysymysten lähettämiseen.

Palvelun julkistuksesta ja siihen liittyvästä tiedottamisesta sovitaan erikseen vastuutahojen kesken.

4. Vuoden 2012 data

Vuoden 2013 aikana portaaliin lisätään ainakin yliopistojen vuoden 2012 julkaisudata, joka kerätään yliopistoilta Excel-formaatissa kevään 2013 aikana. CSC huolehtii tietojen yhdistämisestä, ja ne saadaan Julkaisuportaalin käyttöön todennäköisesti edellisvuoden tapaan Excel-tiedostona. Tarkkaa aikataulua ei ole vielä mahdollista arvioida, mutta on kuitenkin toivottavaa, että tiedot saadaan mahdollisimman lopullisessa muodossa vasta mahdollisten tarkistuskierrosten jälkeen.

Todennäköisesti vuoden 2012 tietojen julkaiseminen tapahtuu aikaisintaan kesällä 2013 ja viimeistään seuraavaan syksyyn mennessä.

4.1. Vuoden 2012 julkaisutietojen lisääminen Julkaisuportaaliin

Jos uudet julkaisutiedot saadaan samassa tai samantapaisessa formaatissa kuin edellisvuonna, niiden konvertointiin käytettävien skriptien muokkaaminen vaatii suhteellisen vähän työtä. Vuoden 2012 dataa täytyy joka tapauksessa muokata samaan tapaan kuin aiemmin, jotta se saadaan mahdollisimman yhdenmukaiseksi ja siinä mahdollisesti esiintyviä virheitä voidaan korjata. Datan yhdenmukaisuuden takaaminen saattaa edellyttää myös aiemman vuoden 2011 datan jonkinasteista muokkaamista.

Portaalin päivitys on mahdollista toteuttaa joko siten, että uudet tietueet ajetaan vanhaan VuFind-instanssiin, tai siten, että portaalia varten perustetaan uusi identtinen instanssi, johon ajetaan molempien vuosien tiedot, ja joka vaihdetaan nimipalvelinmuutoksen avulla vanhan instanssin tilalle sitten kun se on valmis.

4.2. Uudet sektorit

Julkaisuportaaliin on mahdollista liittää yliopistojen lisäksi myös muiden sektoreiden julkaisutietoja, jos ne saadaan kerättyä yliopistojen julkaisutietojen kanssa yhteismitallisessa formaatissa. Jotta kukin julkaisu olisi järjestelmässä vain kerran, sektorirajat ylittävät yhteisjulkaisut tunnistetaan ennen niiden ajamista järjestelmään. Ammattikorkeakoulujen vuoden 2012 julkaisujen viitetiedot kerätään ensimmäistä kertaa OKM:n käyttöön, eli ne olisi periaatteessa mahdollista lisätä portaaliin.

4.3. Uudet julkaisutyypit

Julkaisuportaalin sisältämä data kattaa tällä hetkellä OKM:n tiedonkeruun julkaisutyypit A-E. Näiden lisäksi etenkin yliopiston henkilökunnan julkaisemat väitöskirjat soveltuisivat luontevasti muiden julkaisutietojen yhteyteen, sillä ne ovat merkittävä julkaisutyyppi niillä tieteenaloilla, joilla monografioilla on keskeinen rooli julkaisutoiminnassa.

5. Mahdolliset muut kehittämissuunnitelmat

Vuonna 2013 tehtävän työn lisäksi on käyty keskustelua myös muista pidemmällä tähtäimellä toteutettavista julkaisutietoihin liittyvien prosessien kehittämismahdollisuuksista. Näiden projektien edellyttämää työmäärää tai todennäköistä aikataulua on toistaiseksi vaikea arvioida.

5.1. Tietojen dynaamisempi päivittyminen

Excel-formaatissa toteutetut tiedonsiirrot rajoittavat tällä hetkellä julkaisutietoihin liittyvien prosessien kehittämistä. Käytännössä tietojen siirtäminen eräajoina on ainoa realistinen vaihtoehto. VuFind-ohjelmisto tukee kuitenkin myös dynaamisempaa tietojen importointia (sekä uusien tietueiden luomista että vanhojen tietueiden päivittämistä) teknisten rajapintojen kautta. Järjestelmää olisi siis periaatteessa mahdollista päivittää reaaliajassa joko OKM:n tietovarastosta tai jopa yliopistojen julkaisurekistereistä haravoiduilla tiedoilla.

Reaaliaikaisuus - tai ylipäänsä dynaamisempien prosessien kehittäminen - edellyttää kuitenkin prosessien ja rajapintojen huolellista suunnittelua. Tietojen konvertointi formaatista toiseen täytyisi suorittaa kokonaan automaattisesti. Lisäksi samaa julkaisua koskevien tietojen yhdistäminen ja epätäydellisten tietueiden käsittely aiheuttaisivat merkittäviä haasteita. Todennäköisesti yliopistojen

järjestelmistä kerättävät tiedot olisi joka tapauksessaärkevintä kierrättää ensin OKM:n tietovaraston kautta, jos sen yhteyteen ollaan rakentamassa työkaluja, joilla tietoja voidaan yhdistää ja normalisoida automaattisesti.

5.2. Tekijöiden, lehtien/sarjojen ja kustantajien tunnistaminen

Yliopistoilta kerättävä julkaisudata ei tällä hetkellä sisällä tekijöiden tunnistetietoja, eli samannimisten tekijöiden erottaminen toisistaan tai eri nimimuodoilla esiintyvien tekijöiden tietojen yhdistäminen on käytännössä mahdotonta. Jos tekijöiden tunnistaminen kuitenkin toteutetaan kansallisella tasolla (tutkijarekisteri) tai jos data sisältää muita tekijöiden identifiointiin tarkoitettuja tunnisteita (esim. ORCID tai ISNI), tekijöiden erottelu toisistaan on mahdollista myös VuFindissa. Tämän ominaisuuden käyttöönotto edellyttää kuitenkin suunnittelu- ja kehitystyötä. Lisäksi ongelmia aiheutuu siitä, että vanhimmassa datassa tekijöiden tunnistamista ei ole vielä tehty, eli nämä julkaisut eivät välttämättä yhdisty oikeaan henkilöön tai saattavat yhdistyä väärin.

Julkaisufoorumitietojen hyödyntämisen kannalta on myös tärkeää, että julkaisuihin liittyvät julkaisukanavat - lehdet, sarjat, kustantajat - on tunnistettu mahdollisimman luotettavasti. Lehtien ja sarjojen osalta tunnistamista auttaa oleellisesti, jos tietueet sisältävät niiden ISSN-tunnuksen. Tulevaisuudessa on joka tapauksessa syytä pyrkiä siihen, että tietueisiin lisätään julkaisukanavien tunnistamisen ja julkaisufoorumiluokitusten määrittelyn yhteydessä niistä mahdollisesti puuttuvat ISSN-tunnisteet ja myös lehtien, sarjojen ja etenkin kustantajien nimet normalisoidaan mahdollisuuksien mukaan yhdenmukaiseen muotoon. Tämä parantaisi oleellisesti portaalin haku- ja selailutoimintojen luotettavuutta julkaisukanavia koskevien tietojen osalta.

Lisäksi olisi mielekästä selvittää sitä, millä tavoin Julkaisufoorumia varten toteuttava julkaisukanavalista voitaisiin teknisesti yhdistää Julkaisuportaaliiin - sekä niin että julkaisukanavalistasta on linkkejä Julkaisuportaalissa oleviin julkaisuihin että niin, että joko Julkaisuportaalissa tai ainakin julkaisutietojen muokkaamisessa hyödynnetään tavalla tai toisella lehtilistan sisältämää dataa.

5.3. Julkaisutietojen välittäminen avoimena datana

Julkaisuportaaali tarjoaa käyttöliittymän yliopistojen julkaisutietoihin, mutta jos julkaisutiedot julkaistaisiin soveltuvalla lisenssillä avoimena datana, niille saattaisi olla muutakin käyttöä, ja tutkijoilla olisi mahdollista tehdä tiedoista omia analyysejään. Dataan liittyvistä oikeuksista täytyy joka tapauksessa sopia yliopistojen kanssa.

Datan tallennukseen ja välittämiseen liittyvien järjestelmien ja teknisten rajapintojen osalta keskeinen päätettävä asia on se, mistä järjestelmästä ja missä formaatissa tiedot olisi mahdollista haravoida ulkopuoliseen käyttöön. Vaikka myös VuFindissa on datan haravointiin soveltuvia rajapintoja, tietojen imurointimahdollisuus saattaa kuitenkin olla mielekkäämpää sijoittaa jonkin muun järjestelmän yhteyteen. Tämä on erityisen perusteltua siinä tapauksessa, jos julkaisutietojen lisäksi haravoitavaksi halutaan antaa muitakin tutkimuksenhallinnan tarpeisiin kerättyä dataa (esim. tutkimuksenhallintajärjestelmiä varten kehitetyssä CERIF-formaatissa).